

Algoritmo para el seguimiento de objetos en secuencias de imágenes

PROCESAMIENTO DE IMAGENES

J. J. Aranda

Instituto Central de Investigaciones Digitales (ICID), Ciudad de La Habana, Cuba

Se presenta un algoritmo para el seguimiento de objetos de mediano nivel de segmentación, que se basa en evaluar heurísticas simples sobre el movimiento de los mismos, tales como masa, pequeños cambios en el movimiento y que sea en común, proximidad, brillantez y dirección del movimiento, como aspectos significativos, con lo que se obtiene un conjunto de características de este objeto y se realiza la medición del acople del mismo entre el cuadro actual y el anterior. Este algoritmo es capaz de mantener la localización del objeto a lo largo de una secuencia de imágenes en tiempo real.

In this paper an algorithm with a middle level of segmentation is introduced for object-tracking. It is based on the evaluation of simple heuristics on motion, such as mass, little changes and common motion, neighborhood, brightness and direction of the motion, as significant feature changes. Thus a series of characteristics can be obtained for the object, which permit to track it along a sequence of real-time frames.

Recibido: octubre de 1995

Introducción

El estudio de una secuencia de imágenes es una de las aplicaciones que más interés despierta en la actualidad, debido a sus aplicaciones en la medicina -para el análisis cuantitativo del funcionamiento de los órganos (por ejemplo, valoración de la contractilidad segmentaria (*wall stress*) del ventrículo izquierdo del corazón)- y la robótica -para determinación y comprobación de la trayectoria-, etcétera.

La restricción más fuerte al trabajo la impone la necesidad de trabajar en tiempo real, lo que equivale a no perder la con-

tinuidad del movimiento. Para que este sea perfectamente reconstruible a los ojos humanos, se necesita que el programa resuelva el acople del objeto a través de los diferentes cuadros en un tiempo menor de 83 ms, o sea, que se trabaje al menos a una frecuencia de 12 cuadros por segundo.

El cine y la televisión, para garantizar una mayor calidad, trabajan con un número de cuadros, que es por lo menos el doble de esta frecuencia (24 el cine, 25 normas PAL y CECAM de TV y 30 norma

NTSC), con lo que cumplen los requisitos del teorema de Nyquist.

De acuerdo con esta premisa, cualquier algoritmo previsto para seguir el movimiento de objetos en tiempo real debe ser capaz de trabajar en un tiempo no mayor de estos 83 ms.

En Ballard,¹ se plantean cinco heurísticas para acóplar puntos de imágenes separadas por un pequeño intervalo de tiempo:

1. Velocidad máxima: Si se conoce que un punto tiene una velocidad máxima (V) con respecto a un equipo de captación de la imagen estacionario, solo puede moverse ($V dt$) entre dos imágenes separadas por un intervalo de tiempo (dt). Limita donde puede aparecer este punto en la próxima imagen.

2. Pequeño cambio de velocidad: De acuerdo con las leyes de la Física, los cuerpos tienden a conservar la velocidad (magnitud y dirección) a la que se mueven.

3. Movimiento común: Los objetos espacialmente coherentes aparecen en imágenes sucesivas como regiones de puntos conservando un movimiento común.

4. Acople consistente: Dos puntos de una imagen generalmente no deben acoplar con uno solo de otra. Las oclusiones pueden causar excepciones a esta heurística.

5. Movimiento conocido: Si existe un modelo que supla información sobre los movimientos de los objetos, pueden derivarse, predecirse y reconocerse movimientos retinales.

En el trabajo de Johnson² se plantean tres heurísticas que permiten reducir la región de interés (ROI) o área de búsqueda de un objeto dentro de la imagen:

(a) Proximidad: Si aparece un objeto O_B en el lugar B del cuadro actual I_t que está en proximidad cercana a un objeto en movimiento O_A en el lugar A del cuadro anterior I_{t-2} , hay una elevada probabilidad de que O_B sea O_A , por lo que basta examinar algunas características de O_B para

concluir que es O_A en definitiva. Puede actualizarse la nueva localización y dirección del movimiento de O_A .

(b) Dirección del movimiento: Si la dirección del movimiento desde O_A hasta O_B es cercana a la dirección previa del movimiento hasta la localización A de O_A , hay elevada expectación de que O_B sea O_A .

(c) Brillantez: Si la brillantez de O_B es similar a la de O_A , mientras más cercanas, mayor probabilidad de que sean el mismo objeto.

Se han propuesto muchos métodos para el análisis de las secuencias de imágenes que pueden clasificarse en:³

1. Diferencias espacio-temporales: El desplazamiento se deriva de resolver una ecuación diferencial construida con las derivadas espacial y temporal de la imagen, asumiendo que la superficie de luminancia es localmente lineal. Deben imponerse restricciones adecuadas en el campo de vectores de flujo.

2. Acople de características: Calcula el desplazamiento de objetos buscando el mejor acople entre elementos característicos en dos cuadros consecutivos.

3. Entendimiento: A partir de las relaciones físicas y estructurales de los objetos, inferir su desplazamiento.

Las principales técnicas que se han elaborado para resolver el problema de la correspondencia entre objetos en imágenes diferentes pueden dividirse en tres categorías,⁴ de acuerdo con el nivel al cual establecen la correspondencia y a la cantidad de procesamiento que debe preceder a esta etapa en particular:

A. Enfoque de bajo nivel: Basado en la idea de que la percepción humana depende de la correlación de puntos de similar intensidad y es independiente de la percepción de la forma. Los elementos básicos de correspondencia deben ser los píxeles, no teniendo necesidad de procesarse la imagen previamente al establecimiento de esa correspondencia.

B. Enfoque de medio nivel: La solución del problema del mejor acople se intenta inmediatamente después de la organización de la información que trae la imagen en unidades significativas, usando solo operaciones en tonos de gris que son similares a los del sistema de visión humano. Este enfoque cae entre una distribución espacial de los tonos de gris y el reconocimiento de objetos previo al acople.

C. Enfoque de alto nivel: Tiene sus raíces en el reconocimiento de patrones. El problema muchos objetos-muchos patrones puede adaptarse a: un cuadro representa la colección de patrones y el otro un conjunto de objetos a ser acoplados. Debe disponerse de una elaborada descripción de unidades en las dos imágenes, mucho más compactada que sus píxeles, previa a la etapa de acoplamiento. El razonamiento es que si es utilizable una descripción detallada de formas y otros atributos geométricos de los objetos, es fácil establecer la correspondencia.

En esencia, puede considerarse que cualquier solución al problema del seguimiento de un objeto en una secuencia de imágenes está inmersa en calcular una medida de similaridad capaz de incorporar los aspectos relevantes. Deben realizarse siempre dos acciones:

(a) Transformar los objetos con vistas a dimensionalizar su similaridad.

(b) Realizar la medición actual.

Cualquier intento de cuantificar la similaridad debe obedecer varias restricciones para liberarse de particularidades irrelevantes, garantizando compatibilidad bajo traslación, cambio de escala y, aunque menos fuertemente, rotación.

Descripción del algoritmo

El algoritmo propuesto se fundamenta en las heurísticas expuestas en la sección I: evalúa pequeños cambios en el movimiento y el que este sea común, aprovechando eficientemente las de proximidad, dirección del movimiento y brillantez, lo que le permite eliminar grandes zonas de la imagen al trabajar con ROI pequeñas y en una

vecindad cercana a la que ocupaba el objeto en el cuadro anterior.

Por el método utilizado para su trabajo, pertenece al grupo de acople de características, y por las técnicas utilizadas, se puede afirmar que es de medio nivel, ya que no es necesario terminar el proceso de segmentación para formar el vector de características en la imagen actual y realizar su comprobación contra la imagen anterior por la medición de la distancia entre ambas.

También es posible predecir la ubicación del objeto en el próximo cuadro.

Como características se seleccionan:

o Area: Total de píxeles que pertenecen al objeto.

o Momentos centrales de primer orden:

$$m_{10} \text{ y } m_{01} \quad \dots (1)$$

donde:

$$m_{ij} = \sum_x \sum_y x^i y^j I(x, y),$$

siendo:

$$m_{00}: \text{Masa total de } I.$$

o Coordenadas del centroide:

$$x_0 = m_{10} / m_{00}; \quad y_0 = m_{01} / m_{00} \quad \dots (2)$$

o Invariantes (bajo traslación, rotación y cambio de escala en dos dimensiones): Pueden aplicarse para identificar objetos independientemente de su localización.

$$\text{Inv}_1 = (m_{20} + m_{02}) / m_{00}^2 \quad \dots (3)$$

$$\text{Inv}_2 = [(m_{20} - m_{02})^2 + 4 m_{11}^2] / m_{00}^4 \quad \dots (4)$$

$$\text{Inv}_3 = [(m_{30} + m_{12})^2 + (m_{21} + m_{03})^2] / m_{00}^5 \quad \dots (5)$$

Tono de gris en el centro: Se obtiene como el valor de gris de la función $I(x_0, y_0)$, a partir de las coordenadas del centroide.

Este primer grupo de características son propias del objeto en cada cuadro y sirven para establecer una medida que elimina posibles errores o confusiones por la presencia de otros objetos cercanos e inclusive, por oclusiones parciales.

El otro grupo de características⁶ está relacionado con el movimiento del objeto en sí, por lo que se toma información de dos imágenes I' (cuadro anterior) e I'' (cuadro actual):

$$\text{o Factor de escala: } k = \sqrt{\frac{m''_{00}}{m'_{00}}} \quad \dots (6)$$

o Angulo de rotación:

$$\tau = 0,5 \operatorname{atan} [2m''_{11} (m'_{20} - m'_{02}) - 2m'_{11} (m''_{20} - m''_{02})] / [4 m''_{11} m'_{11} + (m'_{20} - m'_{02}) (m''_{20} - m''_{02})] \quad \dots (7)$$

o Vectores de desplazamiento con respecto al cuadro anterior:

$$px = m''_{00} / (km'_{00}) - \cos \tau x'_0 + \sin \tau y'_0 \quad \dots (8a)$$

$$py = m''_{00} / (km'_{00}) - \cos \tau y'_0 + \sin \tau x'_0 \quad \dots (8b)$$

Con este vector se pueden medir las heurísticas descritas en la sección I.

Como criterio de la distancia entre vectores válida para el acople, se calcula la distancia entre cuadras (*city block*), lo que obedece en primera instancia a que se trata de evitar calcular más raíces cuadradas, para mantener la ejecución en tiempo real:

$$\text{distancia}_i = |\text{Característica}_i - \text{Característica}_i| \quad \dots (9)$$

para i : Elemento del conjunto de las características descritas anteriormente.

Como medida total del acople se utiliza un vector umbral obtenido a partir del análisis del intervalo de variabilidad permisible para cada característica.

Resultados

Para las pruebas se utilizó:

1. Computadora compatible IBM 386 DX 33 MHz.

2. Tarjeta digitalizadora de imágenes de 256x243 con 256 tonos de gris, que permite capturar cuadros a una frecuencia de 60/s.

Con estos medios, se trabajó en el estudio de secuencias de imágenes de objetos del mundo real, utilizando una cámara PANASONIC CCTV modelo WV-1410 con un lente JVC HZ-C611AF(U) que tiene zoom manual con posibilidades de acercamiento desde $+\infty$ hasta 1,2 m con foco automático, lo que permitió simular los movimientos de acercamiento y alejamiento entre la cámara y el objeto, como si fuera un robot, a diferentes velocidades. También se realizaron pruebas de movimiento en *panning* en las direcciones horizontal y vertical, así como combinaciones de todos estos movimientos.

Se eligió una ROI centrada alrededor del punto (x_0, y_0) de 20x20 pixeles, ya que este tamaño contiene en todos los casos a los objetos estudiados.

La secuencia de imágenes de trabajo se estableció en 1 000 como máximo, aunque podía ser menor en el caso, en que el programa perdiera el acople de los vectores de características entre dos cuadros consecutivos, al ser la medida total del acople mayor que la permitida, cosa que no ocurrió en ningún caso.

Los tiempos de ejecución del algoritmo fueron siempre del orden de 33 ms, o sea, que se logró trabajar a la frecuencia de la norma NTSC de televisión: 30 cuadros por segundo.

Conclusiones

Se puede afirmar que se ha desarrollado un algoritmo capaz de mantener el seguimiento de objetos en tiempo real sobre la base de las heurísticas mencionadas de proximidad, dirección del movimiento, brillantez, masa y momentos inerciales que utilizando un enfoque de medio nivel de segmentación es capaz de formar un vector de características con el mínimo de redun-

dancia y garantizar la fidelidad en la descripción del objeto y de su movimiento.

Este resultado tiene aplicaciones importantes en los campos de la medicina y la robótica.

Referencias

1. BALLARD, D. & C. M. BROWN: *Computer Vision*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.
2. JOHNSTON, V. S.; P. DEREK & P. D. LOPEZ: "A Biologically Based Algorithm for Rapid Scene Analysis", *MCCS-88-114* Computing Research Laboratory, Río Grande Research Corridor, New Mexico State University, 1988.
3. GUIDUCCI, A. & P. GRATTONI: "A Contour Segmentation Method for Dynamic Image Analysis", *Time-Varying Image Processing and Moving Object Recognition*, 2nd ed., Cappellini, pp. 331-338, Elsevier Science Publishers B. V., 1990.
4. DASKALAKIS, T. N. & C. N. DASKALAKIS: "Knapsack Formulation of Image Matching", *ICASSP 91*, Vol. 4, pp. 2365-2368, CH2977-7/91/0000-2365, IEEE, 1991.
5. ROSENFELD, A. & A. C. KAK: *Digital Picture Processing*, 2nd ed., Academic Press, Inc., 1982.
6. SLUZEK, A.: "Real Time Determining of Position and Orientation of Moving 3D Objects", *Time-Varying Image Processing and Moving Object Recognition*, 2nd ed., V. Cappellini, pp. 272-279, Elsevier Science Publishers B. V., 1990.

JUAN JOSE ARANDA ABOY, Licenciado en Ciencias de la Computación, Investigador Auxiliar, Profesor Auxiliar Adjunto, Trabaja en investigaciones sobre detección automática de contornos en imágenes médicas